Minería de datos y Patrones

Version 2024-1

Feature Selection
Clean

[Capítulo 3]

Dr. José Ramón Iglesias
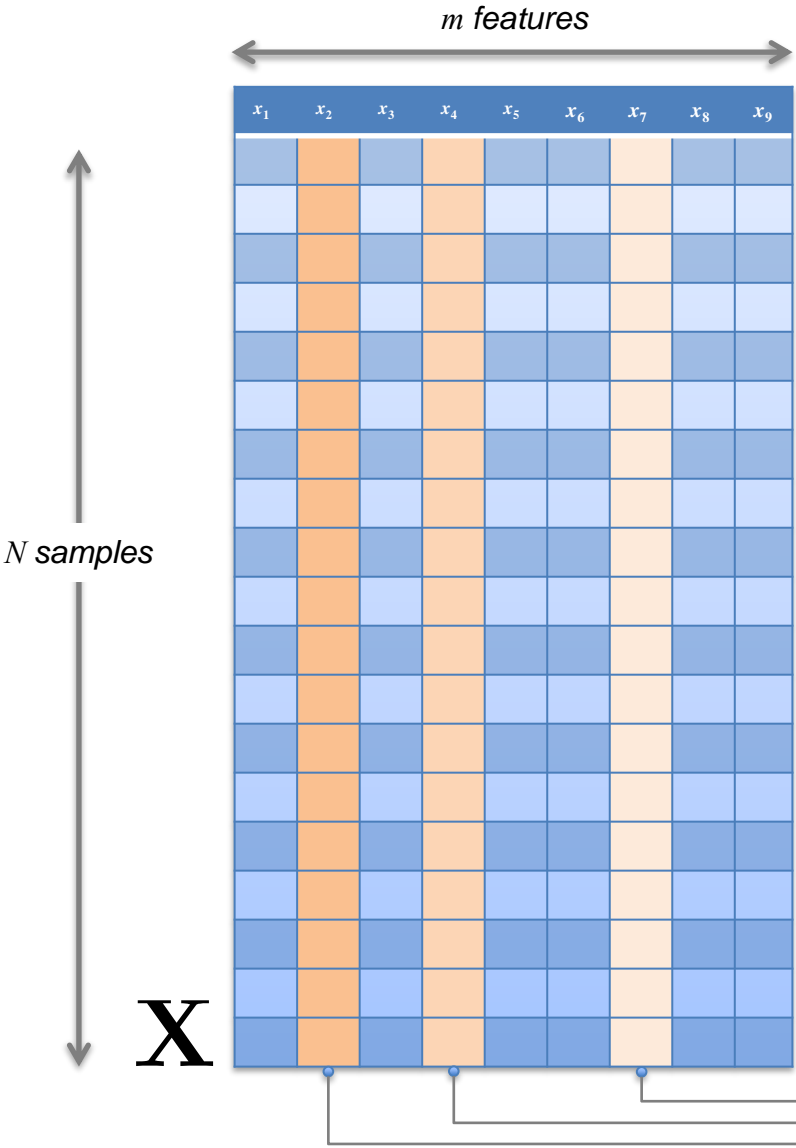
DSP-ASIC BUILDER GROUP

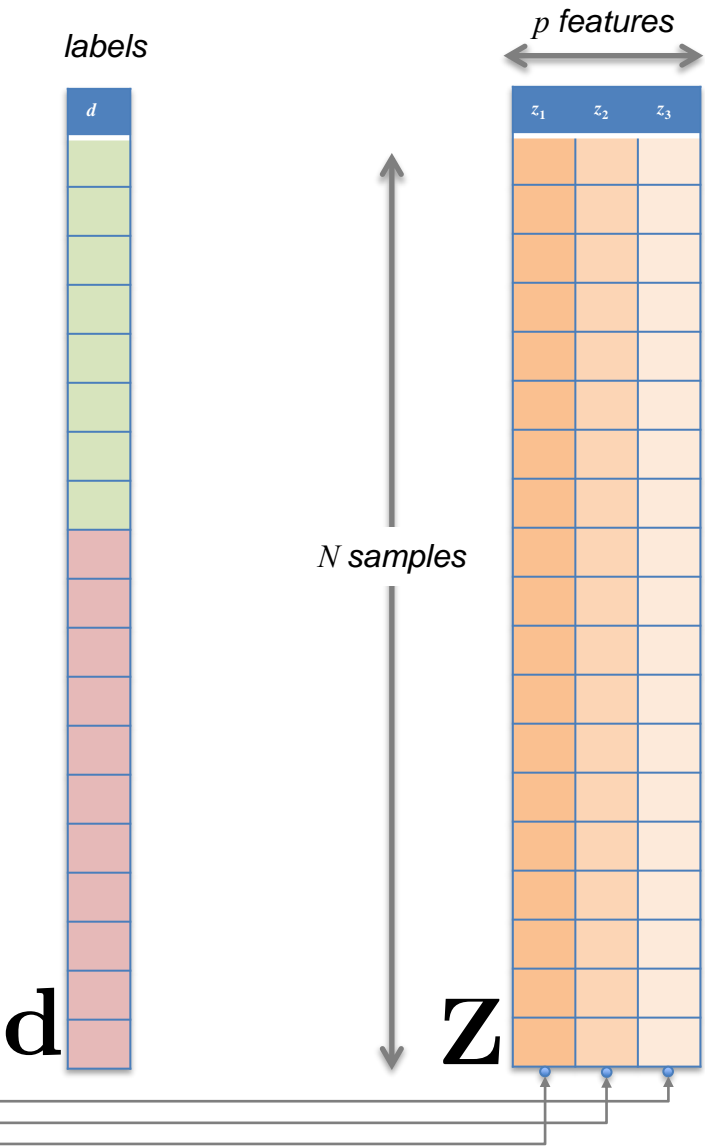Director Semillero TRIAC

Ingenieria Electronica

Universidad Popular del Cesar

Extracted features

m features

x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9

N samples

X

labels

d

d

Selected features

p features

z_1 z_2 z_3

N samples

Z

Clean Algorithm

(this step must be done before any feature selection algorithm)

1. Compute correlation between columns of X
2. If there are columns highly correlated select one randomly and remove the rest
3. Eliminate constant features (std < 10^{-8})